DOCUMENT RESUME

ED 048 333 TM 000 393

AUTHOR	Stufflebeam, Daniel L.
TITLE	Evaluation as Enlightenment for Decision-Making.
INSTITUTION	Ohio State Univ., Columbus. Evaluation Center.
PUB DATE	19 Jan 68
NOTE	52p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Decision Making, Educational Accountability, Educational Change, *Evaluation, Evaluation Criteria, *Evaluation Needs, *Evaluation Techniques, Guides, Models, Program Evaluation, *Research Design, Research Problems
IDENTIFIERS	*Elementary Secondary Education Act Title I, Title III

ABSTRACT

The need for competent formal evaluation programs, particularly for new federally assisted programs, is expressed. Problems in defining educational evaluation and its requirements, in designing such evaluations, and 3 possible sources of faulty conceptual bases for evaluations, are presented. An attempt is made to define evaluation in general, to analyze emergent problems of educational change, and to identify the types of decisions for which evaluations are needed in these programs. Four strategies for evaluating educational programs are outlined. These include context, input, process, and product evaluation, each of which are used at a distinct stage in the development of a program. Finally, a general guide for developing evaluation designs to implement a given evaluation program is provided. The logical structure of evaluation design is presented in a step by step format. (PR)

TM

EDO 48333

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

10. 13. 70

EVALUATION AS ENLIGHTENMENT FOR DECISION-MAKING

Daniel L. Stufflebeam

An Address Delivered at the
Working Conference on Assessment Theory
Sponsored by
The Commission on Assessment of Educational Outcomes
The Association for Supervision and Curriculum Development
Sarasota, Florida
January 19, 1968

EVALUATION CENTER
THE OHIO STATE UNIVERSITY
College of Education

EA 002 367

TM 000 393 1

The EVALUATION CENTER, an agency of the College of Education, is committed to advancing the science and practice of educational evaluation. More specifically, the purpose of the Center is to increase education's capability to obtain and use information for planning, programming, implementing and evolving educational activities. To serve this purpose, the Center's interdisciplinary team engages in research, development, instruction, leadership and service activities.

HISTORY

The origin of the present Center traces back to the establishment of the Ohio State University Test Development Center in 1962. Due to the urgent need for a more comprehensive approach to evaluation than that afforded by standardized testing, the Test Development Center was expanded in 1965 into the present Evaluation Center which is concerned with many modes of evaluation in addition to standardized testing. However, test development remains an important part of the Evaluation Center program.

GOALS

The broad objectives of the currently constituted Center are:

to increase scientific knowledge of educational evaluation and planning;
to develop evaluation strategies and designs;
to develop evaluation methods and materials;
to provide instruction in evaluation;
to disseminate information related to educational evaluation;
to assist educationists in evaluating their programs.

THE OHIO STATE UNIVERSITY

EVALUATION CENTER

1712 NEIL AVENUE
COLUMBUS, OHIO 43210

ORGANIZATION

To serve its complex objectives, the Center has developed an interdisciplinary team. Currently, the staff of the Center consists of fifty-four members, including five professorial positions, plus a varying number of visiting faculty. The staff and visiting professors bring expertise from the fields of economics, education (administration, curriculum and supervision, elementary and secondary school teaching, evaluation, mathematics, planning, research methodology, and tests and measurement), psychology, sociology, systems analysis, and urban planning. The Center is organized into four divisions: Administration and Program Development; Leadership in Evaluation; Research in Evaluation; and Test Development. The Center is administered by a director and an associate director for each division.

STAFF

Daniel L. Stufflebeam, Director
Michael S. Caldwell, Associate Director
Administration and Program Development
Edwin G. Novak, Associate Director
Research in Evaluation
Jack M. Ott, Associate Director
Test Development
Blaine R. Worthen, Associate Director
Leadership in Evaluation

ED048333

CONTENTS

	Page
INTRODUCTION	1
PART I: The State of the Art in Educational Evaluation	3
The Setting	3
The Need for Better Educational Evaluations	4
Problems in Educational Evaluation	7
Problems in Defining Requirements for Educational Evaluations	9
Problems in Defining Educational Evaluation	10
Problems in Designing Educational Evaluations	11
PART II: The Nature of Evaluation	17
The General Nature of Evaluation	18
Evaluation Defined	19
Evaluation in Fields Other Than Education	20
Evaluation in Education	22
A Rationale for Educational Evaluation	24
Four strategies for Evaluating Educational Programs	30
Context Evaluation	30
Input Evaluation	33
Process Evaluation	35
Product Evaluation	36
The Structure of Evaluation Design	39
Design Defined	39
Focusing the Evaluation	41
Collection of Information	44
Organization of Information	45
Analysis of Information	46
Reporting of Information	46
Administration of Evaluation	47

INTRODUCTION

Chairman Beatty, and ladies and gentlemen: it is a pleasure to be here; and I appreciate the opportunity to test some of my ideas about educational evaluation before this distinguished group.

For the past two and one-half years I have been heavily engaged in evaluation activities with personnel from local schools, state education departments, and the United States Office of Education. Those activities, for the most part, have involved efforts to evaluate projects funded under Title I and Title III of the Elementary and Secondary Education Act of 1965. This paper is based on those experiences and is an attempt to summarize some of my ideas about the kinds of evaluation which are needed in current programs of educational change.

The paper is divided into two parts. Part I is concerned mainly with determining the present state of the art in educational evaluation. In this part, I have attempted to describe current requirements for educational evaluation, to illustrate that educators have thus far been ineffectual in their attempts to meet these requirements, and to point out some possible reasons for poor evaluations in education. In Part 2 of the paper, I have attempted to conceptualize some alternative approaches to educational evaluation. This second part of the paper includes attempts to define evaluation in general terms, to sketch four evaluation strategies which I think have particular relevance to educational change activities, and to explicate the structure of evaluation design.

Before proceeding to present the body of the paper, I want to emphasize that my formulations are largely untested and are therefore highly

tentative. I sincerely hope that you find these rough ideas worthy of your examination. If you find any of them to be viable, I hope that you will help me, both during and after this working conference, to refine and extend them. Without further introduction, let me proceed with the presentation of Part I.

Part I: The State of the Art in Educational Evaluation

The Setting

Education is becoming increasingly valued as a means to meet the social and economic as well as the intellectual needs of society. To fulfill this expanding role, educators are being asked to deal with many critical societal problems. These include inequality of opportunities among racial groups, de facto segregation, riots in our cities, disillusionment of youth, and school dropouts. Clearly, the rising trend of these problems must be curbed and pushed back for the welfare of our civilization. Education is thus being given a most urgent and difficult charge, and to meet this charge educators must mount many new and innovative efforts.

To help educators meet their new responsibilities, society is annually providing billions of dollars through federal, state and foundation programs to education agencies at all levels. Examples of increased support to education include the Elementary and Secondary Education Act of 1965, the Headstart Program, the Education Professions Act, and the Experienced Teacher Fellowship Program. Many industries are also developing education components, and soon we will probably see many new education-industry combines and consortia. Clearly, in addition to new responsibilities education also has unprecedented opportunities to improve and expand its programs.

These opportunities, however, have also carried requirements that educators evaluate their new plans and programs. These requirements are especially evident in new federal assistance programs, e.g., Title I and Title III of the Elementary and Secondary Education Act. Here, the law explicitly states that fund recipients will make at least annual evaluation reports. As a consequence, many educators at all levels for the first time are having to cope with requirements for formal evaluation.

Such requirements for evaluation seem reasonable; and, in my judgment, they are long overdue. Funding agencies and the public have the right to know whether their huge expenditures for education are producing the desired effects. Even more important than this, educators themselves need evaluative information to provide rational bases for their decisions among alternative plans and procedures. However, to justify requirements for evaluation is not to operationalize them. Educators must repond to the requirements, and they must do so effectively.

The Need for Better Educational Evaluations

Without question, educators are responding to requirements for evaluation. The multitude of evaluation reports now available from local schools, state education departments, regional educational laboratories, etc. demonstrates that educators are expending significant amounts of time, effort, and money to evaluate their programs. However, the increased activity alone has not met the need for

effective evaluations. While educators have been busy doing evaluations, the fruits of their efforts have not provided the information needed to support decision-making related to the programs being evaluated.

Many of the completed evaluation reports contain only impressionistic information. Though such information may be pertinent to the concerns of decision-makers, it usually lacks the level of credibility required by decision-makers to defend their decisions, and seldom can such information be of material use in making important decisions. A case in point is the first annual report for Title I of The Elementary and Secondary Education Act.[1] This report was highly important since it encompassed the thousands of Title I projects throughout the nation. However, it fell far short of being a useful document, for it was almost devoid of hard data. On the other hand, it did contain many anecdotal accounts wherein persons who were responsible for conducting Title I activities stated that they felt that their program was being successful; and many of them speculated as to the reasons for the alleged successes. Though these anecdotes may have touched key issues related to the improvement of the billion dollar per year Title I program, decision-makers in the Congress, the Office of Education, state education departments, and local school districts could hardly base important decisions on a few "possibly accurate" pieces of testimony.

---

[1] Public Law 89-10: The Elementary and Secondary Education Act of 1965, Title I.

The situation is not much different in Title III of the Elementary and econdary Education Act. Title III staff members in the U. S. Office of Education have continuously ranked the quality of Title III projects on a five point scale for each of fifteen criteria.[2] The criterion relating to evaluation has consistently been ranked near the "poor" end of the scale and lower than thirteen of the other criteria--the exception being the criterion related to dissemination. Guba has also suggested that evaluation plans in Title III projects are weak.[3] Based on his analysis of thirty-two Title III projects, Guba concluded that "It is very dubious whether the results of these evaluations will be of much use to anyone. They are likely to fit well, however, into the conventional school man's stereotype of what evaluation is: something required from on high that taked time and pain to produce but which has very little significance for action."[4]

Unlike the Title I and Title III evaluations referenced above, some evaluations provide for hard data. For example, the evaluation report for New York City's Higher Horizons Program[5] used rigorous

---

[2]These criteria are listed on pp. 70-71 of the current Title III guidelines. (A Manual for Project Applicants and Grantees, Washington, D. C.: U. S. Office of Education, 1967.)

[3]Egon G. Guba. Evaluation and the Process of Change, Notes and Working Papers Concerning the Administration of Programs authorized under Title III of Public Law 89-10, The Elementary and Secondary Education Act of 1965 as amended by Public Law 89-750, April 1967, p. 312.

[4]Ibid

[5]Wayne J. Wrightstone, et al. Evaluation of the Higher Horizons Program for Underprivileged Children, Cooperative Research Project No. 1124, Bureau of Educational Research, Board of Education of the City of New York.

research procedures to compare the performance of an experimental group receiving the Higher Horizons Program with the performance of a control group which was matched to the experimental group on several counts. The basic conclusions contained in this nearly 300 page report were typical of findings for rigorous educational evaluations: "there were no significant differences." In sharp contrast, however, the report also noted that the teachers and principals who had been involved in the program said that it was making differences so significant that the program simply could not be abandoned.

Though the Title I, Title III and Higher Horizons evaluations differed as to rigor, they were alike in one respect. None of them provided much help to the decision-maker for improving the programs being evaluated. While I have cited only three examples of the deficiencies in current evaluations, I think they are sufficiently weighty ones to illustrate my point. In too many cases, evaluation reports provide little or no help to decision-makers, and decision-making in and about education must remain an arty endeavor.

Problems in Educational Evaluation

What is the explanation for this situation? Why is it that educators are failing to provide evaluations which are at the same time useful and scientifically respectable? Why is it that evaluations which adhear to classical research methods provide information which is of only limited help in making decisions about programs, and why do the typical "no significant difference" findings in so many of these evaluations contravene the experiences of those who are intimately involved in the programs?

One cannot answer these questions simply on the grounds that evalua-

tion practice lags too far behind evaluation theory, or that there is a lack of effort on the part of educators to evaluate their programs. Further, it is not enough to note that evaluation testimony given by witnesses is not credible, or that typical findings of no significant differences are correct because nothing in education ever makes a difference. Rather, I think the lack of adequate evaluation information persists because of several fundamental problems which must be solved before educators can improve their evaluations. These include a lack of trained evaluators, a lack of appropriate evaluation instruments and procedures, and a lack of adequate evaluation theory. In my judgment, the most basic of these problems is a lack of adequate theory or conceptualizations pertaining to the nature of evaluations which are needed to accommodate educational programs.

Clearly, the conceptual bases for evaluations are of fundamental importance. If these conceptions are faulty, then the evaluations which are based on them must also be faulty. Thus, it would seem highly important to identify and examine the efficacy of conceptualizations which underlie current needs for evaluation as well as educators' attempts to meet these needs. It will be useful to divide these conceptualizations into three classes and to consider each one separately. The three classes are:

1. Conceptions of the nature of the educational programs for which evaluations are needed;
2. Conceptions of the nature of evaluation, in general, and as related to specific classes of educational programs; and
3. Conceptions of the structure of evaluation designs needed to conduct educational evaluations.

Problems in Defining Requirements for Educational Evaluations

First, let us examine problems involved in providing an adequate focus for educational evaluation studies. Obviously, to evaluate, one must know what is to be evaluated. Gaining knowledge of what is to be evaluated, however, is currently a difficult task at best. Current needs for educational evaluation have arisen due to programs and activities which are new to the field of education. Such activities involved responsibilities newly assigned to educators, new kinds of relationships among different kinds and levels of agencies, and a need for cooperative decision-making about education among a variety of education and non-education agencies. It should come as no shock if the evaluation theory which has traditionally been viewed as appropriate for education is found no longer to be adequate to meet the information requirements in new educational programs. Clearly, many of the new programs in education are dramatically different from those of the past; and our evaluations should probably be geared to answer questions which are much different from those they have answered in the past.

What we need, I think, are conceptualizations to account for decision processes and information requirements in new educational programs. Programs to improve education depend heavily upon a variety of decisions, and a variety of information is needed to make and support those decisions. Evaluators charged with providing this information must have adequate knowledge about the relevant decision processes and associated information requirements before they can design adequate evaluations. They need to have knowledge about the locus, focus, timing, and criticality of decisions to be served. At present no adequate knowledge of decision processes and

associated information requirements relative to educational programs exists. Nor is there any ongoing program to provide this knowledge. In short, there are no adequate conceptualizations of decisions and associated information requirements or programs to produce them.

Problems in Defining Educational Evaluation

Next, let us attend to problems pertaining to the meaning of educational evaluation. Usually educators have defined evaluation as the science of determining the extent to which objectives have been achieved. The first step in operationalizing this definition is to state program objectives in behavioral terms. Then one must define and operationalize criteria for use in relating outcomes to the objectives. Operationalizing such criteria includes the specification of instruments for measuring outcomes and standards for use in assigning values to the measured outcomes. Standards may be either in absolute or relative terms. An absolute standard might be that students on the average should achieve at least some specified score on a selected achievement test. A relative standard might be that the group of students receiving a new program should achieve scores on a selected achievement test which on the average are higher than scores achieved by an equivalent group of students which received some alternative program. Regardless of the type of evaluative standard used, the data from such studies are analyzed after a complete cycle of the program to determine the extent to which the objectives were achieved.

Evaluations based upon the above definition of evaluation yield data about gross total program effects and then only in retrospect. Such data are useful for making judgments about a project after it has run full

cycle, but they certainly are not adequate to assist educators in the initial planning and in the actual carrying through of programs. At best, therefore, such evaluations provide an insufficient solution to the evaluation problems of educators who must plan and execute innovative programs.

The inadequacy of extant conceptions of evaluation is illustrated by the following excerpt from testimony pertaining to Title I evaluations given before a Congressional committee by a citizens' group in New York City:

> We ask for amendments to render the required evaluations of Title I projects meaningful. The Act states that evaluations must be made, not that they be utilized in future planning. In New York City this year, projects were recycled before last year's evaluations were submitted. To be made more useful, evaluations should have built into them alternatives and the recommendations of the evaluator. What is now an expensive exercise should be made a function to provide service to local school boards having the responsibility for making policy based on experience. American business would not survive if its consultants did not supply management with alternatives after reviewing the efficacy of programs.[6]

Here, the major concern seems to be that reports yielded by current evaluation programs are neither sufficiently specific nor timely to influence educational programs. Obviously, evaluations which do not at least meet these two criteria are of little use.

Problems in Designing Educational Evaluations

Finally, let us consider problems relating to the methodology of evaluation. If current conceptions of evaluation are not adequate for evaluating current educational activities, neither can extant designs be adequate. For, existing means for evaluation have been developed to serve

---

[6] Citizens' Committee for Children of New York, Inc. Newsletter, Statement of Mrs. Nathan W. Levin, Chairman of the Educational Services Section before the Sub-Committee on the Elementary and Secondary Education Act of the Education and Labor Committee of the House of Representatives, March 18, 1967.

the ends of evaluation as they have been conceived traditionally.

The inadequacy of extant evaluation methodology is revealed when one examines the designs educators use to evaluate their programs. If they use a design at all, it typically is an experimental design. The fundamental concern of experimental design is that data which are produced be internally valid, i.e., unequivocal. Several conditions are necessary to meet this criterion. The units to be measured should be randomly assigned to treatment and control conditions. For example, a set of students might be partitioned randomly into two groups--one to receive a new program, the other to receive the school's present offering in the area to be served by the new program. Next, the treatment and control conditions must be applied and held constant throughout the period of the experiment, i.e., they must conform to the initial definitions of these conditions. The new or traditional program conditions could not be modified in process, since in that event one could not tell what was being evaluated. Also, all students in the experiment must receive the same amount of the treatment to which they are assigned; and care must be taken so that students receiving one treatment are not contaminated by the other treatment. If contamination occurred, one could not tell what had caused what after the project was completed. Therefore, until an experiment is completed, one must resist the temptation to apply the successful activities of one condition to students receiving a different condition, even if the activities in the latter condition are obviously failing. Finally, an instrument which is valid and reliable for the specified criterion variable must be administered after a certain period of time--usually a complete program cycle--to subjects from both parts of the experiment. Then, if all of the above condi-

tions were met, one could use predetermined statistical procedures and decision rules to determine unequivocally that there were--or were not significant differences between the experimental and control groups on the outcome variable of interest.

On the surface, the application of experimental design to evaluation problems seems reasonable, since traditionally both experimental research and evaluation have been used to test hypotheses about the effects of treatments. However, there are four distinct problems with this reasoning.

First, the application of experimental design to evaluation problems conflicts with the principle that evaluation should facilitate the continual improvement of a program. Experimental design prevents rather than promotes changes in the treatment because treatments cannot be altered in process if the data about differences between treatments are to be unequivocal. Thus, the treatment must accommodate the evaluation design rather than vice versa; and the experimental design type of evaluation prevents rather than promotes changes in the treatment. It is probably unrealistic to expect directors of innovative projects to accept conditions necessary for applying experimental design. Obviously, they can't constrain their treatment to its original definition just to ensure internally valid end-of-year evaluative data. Rather, project directors must use whatever evidence they can obtain to continually refine and sometimes radically change both the design and its implementation. It is thus contended here that conceptions of evaluation are needed which would result in evaluation programs which would stimulate rather than stifle dynamic development of programs.

A second flaw in the experimental design type of evaluation is that it is useful for making decisions after a project has run full cycle but almost useless as a device for making decisions during the planning and implementation of a project. It provides data after the fact about the relative effectiveness of two or more treatments. Such data, however, are neither sufficiently specific and comprehensive nor are they provided at appropriate times to assist the decision-maker to determine what a project should accomplish, how it should be designed, or whether the project activities should be modified in process. At best, experimental design evaluation reflects post hoc on whether a project did whatever it was supposed to do. At that time, however, it is too late to make decisions about plans and procedures which have already largely determined the success or failure of the project.

Guba[7] has pointed out a third problem with the experimental design type of evaluation; it is well suited to the antiseptic conditions of the laboratory but not the septic conditions of the classroom. The potential confounding variables must either be controlled or eliminated through randomization if the study results are to have internal validity. However, in the typical educational setting this is nearly impossible to achieve. For example, consider the following quotation from an evaluation report completed by Julian Stanley:

---

[7]Egon G. Guba. "Methodological Strategies for Educational Change," Paper presented to the Conference on Strategies for Educational Change, Washington, D. C., November 8-10, 1965.

Even if the program does have considerable cumulative influence on a person's career, this may be slow in appearing and so interactive with other influences that it cannot be discerned clearly by the person himself or by others.

Nevertheless, we must use whatever evidence that can be adduced to determine whether or not such programs are worth repeating and, if so, how they should be modified in order to be more effective. Ideally, in the experimental design sense, we should conduct the program as a controlled experiment, with a well-matched control group that does not attend the institute, and follow up both groups for quite a few years in order to determine how they diverge. If recruiting begins early enough and the applicant group is able enough to provide both groups at a sufficiently high level, this might be done, though the "reactivity" of the disheartened rejectees, the self-fulfilling prophecy of the rejectees, and the inability to control the summer activities of the rejectees might undesirably affect the outcome of the experiment. Merely having on one's record the fact of attending a certain prestigious program, like displaying one's Phi Beta Kappa key, might be a powerful aid...Our chief way of evaluating the success of the program is via reports from staff and participants, particularly the latter.[8]

In the above quotation, Professor Stanley has pointed to many of the reasons why experimental design does not seem well suited to evaluation problems in education. In many innovative programs there clearly are a multitude of confounding factors which simply cannot effectively be controlled.

The existence of potentially confounding factors such as those named by Stanley gives rise to a fourth kind of problem inherent in the experimental design type of evaluation. <u>While internal validity may be gained through the control of extraneous variables, such an achievement is accomplished at the expense of external validity</u>. If the extraneous variables are tightly controlled, one can have much confidence in the findings pertaining to how an innovation operates in a controlled environment. However,

---

[8]Julian Stanley. <u>Benefits of Research Design; A Pilot Study</u>, Final Report, Project No. X-005, Grant OE5-10-272, U. S. Department of Health, Education and Welfare, Office of Education, Bureau of Research, August 1966.

such findings may not at all be generalizable to the real world where the so-called extraneous variables operate freely. Clearly, it is important to know how educational innovations operate under real world conditions.

Thus far, in this paper, I have attempted to depict the state of the art in educational evaluation. To begin with, I pointed out that educators are being faced with many new and different requirements for evaluation. Then I attempted to establish that educators' attempts to meet these requirements thus far have been ineffectual. Finally, I suggested that there are three types of conceptual problems which prevent educators from providing effective evaluations. These are:

1. a lack of understanding of decision processes and information requirements in current programs of educational change;
2. the lack of a definition of educational evaluation which is pertinent to emergent requirements for educational evaluation; and
3. a lack of appropriate evaluation designs.

In the remainder of this paper, I shall attempt a response to these problems by suggesting some alternative conceptions regarding the nature of educational evaluation.

Part II: The Nature of Evaluation

In Part I, I attempted to define some of the current needs and problems in educational evaluation. Since this is a working conference, I should probably stop here so that you could examine my statement of the problem and modify or replace it. After we had achieved agreement as to what the real problems are, we could then proceed to develop relevant solutions. However, I have been asked by the organizers of this conference to expose some of my ideas regarding solutions for current evaluation problems as I see them. As I stated in Part I, I think the basic problem in educational evaluation is a lack of adequate conceptualizations regarding a rationale for and the meaning of evaluation in the context of emergent programs of educational change. Thus, in the remainder of this paper, I shall propose some alternative conceptions regarding the nature of educational evaluation. I am acutely aware, however, of the tentative and untested nature of my formulations. I present these ideas to you in a heuristic spirit in the hope that you will help me examine and refine them.

This part of the paper is divided into four major sections. The first section is an attempt to define evaluation in general. Then, in Section 2, an attempt is made to analyze emergent programs of educational change and to identify the types of decisions for which evaluations are needed in these programs. Section 3 contains outlines of four strategies for evaluating educational programs, and the paper is concluded in Section 4 with an attempt to outline the structure

of evaluation design. To begin, I want to suggest a general rationale for the use of evaluation.

The General Nature of Evaluation

A Rationale

If decision-makers are to make maximum, legitimate use of their opportunities, they must make sound decisions regarding the alternatives available to them. To do this, they must know what alternatives are available and be capable of making sound judgments about the relative merits of the alternatives. This requires relevant information. Decision-makers should, therefore, maintain access to effective means for providing this information. Otherwise, their decisions are likely to be functions of many undesirable elements. Under the best of circumstances, judgmental processes are subject to human bias, prejudice and vested interests. Also, there is frequently a tendency to over-depend upon personal experiences, heresay evidence, and authoritative opinion; and, surely, all too many decisions are due to ignorance that viable alternatives exist. Clearly, the quality of programs depends upon the quality of decisions in and about the programs; the quality of decisions depends upon decision-makers' abilities to identify the alternatives which comprise decision situations and to make sound judgments of those alternatives; making sound judgments requires timely access to valid and reliable information pertaining to the alternatives; and the availability of such information requires systematic means to provide it. The processes necessary for providing this information for decision-making

collectively comprise the concept of evaluation. Given this rationale, I will now attempt to define what I mean by evaluation.

Evaluation Defined

Generally, evaluation means the provision of information through formal means, such as criteria, measurement, and statistics to supply rational bases for making judgments which are inherent in decision situations. To clarify this definition, it will be useful to define several key terms. A decision is a choice among alternatives. A decision situation is a set of alternatives. Judgment is the assignment of values to alternatives. A criterion is a rule by which values are assigned to alternatives, and optimally such a rule includes the specification of variables for measurement and standards for use in judging that which is measured. Statistics is the science of analyzing and interpreting sets of measurements. And, measurement is the assignment of numerals to entities according to rules, and such rules usually include the specification of sample elements, measuring devices and conditions for administering and scoring the measuring devices. Stated simply, evaluation is the science of providing information for decision-making.

The methodology of evaluation includes four functions: collection, organization, analysis, and reporting of information. Criteria for assessing the adequacy of evaluations include validity (is the information what the decision-maker needs?), reliability (is the information reproducible?), timeliness (is the information available when the decision-maker needs it?), pervasiveness (does the information

reach all decision-makers who need it?), and credibility (is the information trusted by the decision-maker and those he must serve?).

Evaluation in Fields Other Than Education

The concept of evaluation as defined above is general, since the assigning of values to alternatives is common to all forms of human thought and activity, and since men have always sought to establish rational defensible bases for their judgments. However, there are many kinds of evaluation which meet the conditions of the above definition, but which nevertheless may be distinguished one from the other. For example, market research, cost benefit analysis, experimental design, objective testing, operational analysis, operations analysis, operations research, Program Evaluation and Review Technique, Program Planning and Budgeting System, quality control, and systems analysis all fit the general definition of evaluation given above. Each of these modes of inquiry is the application of systematic means to aid in the assignment of values to the alternatives in decision situations. These different kinds of evaluation may be differentiated by the decision situations they serve, the settings within which the decisions are made, the kinds of tools and techniques used, the level of precision in the information collection and analytical modes, and the methodological skills of those who conduct the evaluations and those who are served by the evaluations. These substantive and methodological differences probably explain why different names have been given to these different forms of evaluation. For example, consider the following

statement by Quade: "Evaluations undertaken to enable decision-makers to choose among systems, to discover whether a given system would accomplish its objectives, or to set up a framework within which tests of a system could be prepared came naturally to be called 'systems analysis.'"[9] While Quade acknowledged that systems analysis is a form of evaluation, he also noted that the name systems analysis was derived from the nature of this form of evaluation.

Historical review of the more highly developed forms of evaluation listed above reveals that each was developed for relatively specific applications. Program Evaluation and Review Technique was developed to aid the military in making decisions in the development of complex weapon systems. Systems analysis was developed to aid the military in making decisions in the development and implementation of military operations. Experimental design was especially useful for making judgments about the relative merits of agricultural products. And, initially, objective testing was utilized largely as an aid to the military in selecting men for military service. Clearly, the development of each of these forms of evaluation was precipitated by critical decision-making needs; and these forms of evaluation were thus based upon the types of decisions to be served and the settings within which they were to be made. New

---

[9]Edward S. Quade, Editor. Analysis for Military Decisions, Rand McNally and Company, Chicago, 1967, p. 4.

approaches to evaluation were developed because extant approaches did not fit the decision-making requirements as precisely as needed, and because the decisions to be made could have serious consequences if wrong choices were made. Military decisions could effect the outcome of wars; thus, operations research, systems analysis, etc. were developed. Business decisions could result in profit, loss, or bankruptcy for thousands of stockholders; thus, cost-benefit analysis and market research were developed.

Evaluation in Education

In the past, decisions about education have had effects less tangible than those in business, agriculture and the military. Thus, there have not been pressures in education equivalent to those in other fields to motivate the development of highly specialized forms of evaluation to serve well defined classes of educational decisions. Indeed, most educators would be hard pressed to identify and define the critical decision situations in education which merit specialized means for evaluation. It cannot be said, however, that education has been devoid of evaluation practices. Standardized testing has been developed to a high art to aid in college entrance decisions, the passing or failing of students, the assignment of diplomas and degrees, and the placement of students in educational programs. The Buros Mental Measurement Yearbooks[10] have been developed to aid educators in the selection and use of tests. And, recently, Project EPIE (Educa-

---

[10] Oscar K. Buros. The Buros Mental Measurement Yearbooks, Highland Park, New Jersey: The Gryphon Press, 1949.

tional Products Information Exchange)[11] has been developed to assist educators in selecting from among alternative products which are related to education. Generally, however, educators have failed to develop specialized means to aid their decisions about programs.

A prevalent position in education has been to avoid "reinventing the wheel," but instead to look to other fields where problems similar to those in education have been faced and solved. This reasoning has led educators to adopt such evaluation modes as experimental design. Here a technique, previously utilized to assist farmers to select from among alternative kinds of fertilizer and seed, is being used to assist educators to select from among alternative educational innovations. The analogy between educational innovations and fertilizer is hopefully remote. More recent forms of such borrowings are those of Program Evaluation and Review Technique, systems analysis, and the Program Planning and Budgeting System. At this point I would like to note that selective borrowing from other fields can save educators a great deal of time and effort. However, I also want to caution that wholesale, non-selective borrowing of techniques from other fields can result in the misapplication of techniques which never were intended for and do not fit educational situations. I think that educators' use of experimental design to evaluate innovative programs is an example of what can happen in the latter case. The use of experimental design in such applications has cost educators

---

[11] The EPIE Forum, A Monthly Publication of the Educational Products Information Exchange Institute Created by and for Professionals in Education, New York: Educational Products Information Exchange Institute.

much time and effort without yielding much assistance for decision-making.

As stated earlier in the paper, I think educators need some new basic conceptualizations to enable development of evaluation theory and methodology which has specific relevance to educational problems. In the previous section I have suggested a general rationale and definition for evaluation. Now I will attempt to derive a rationale and definition for evaluations in education.

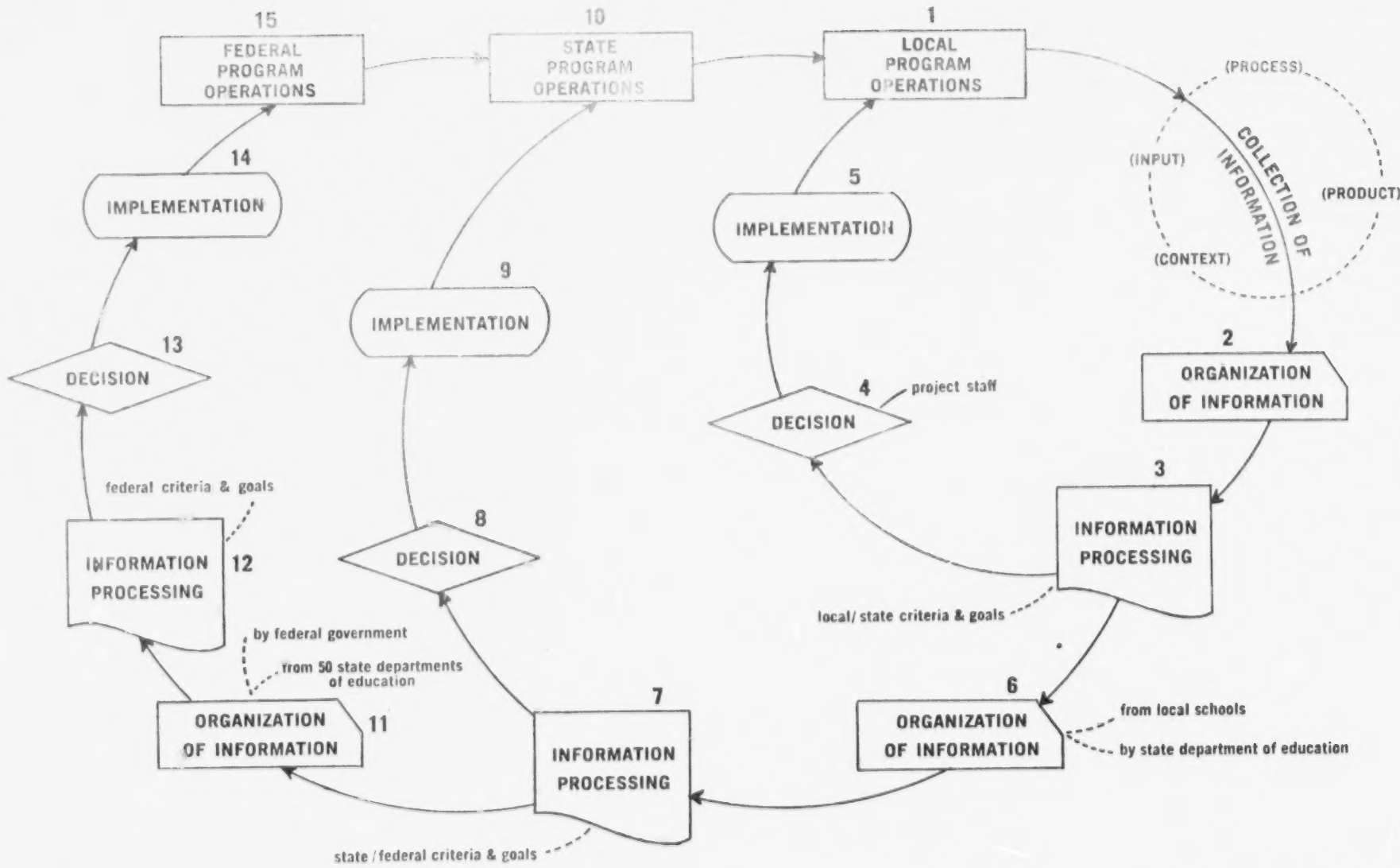
A Rationale for Educational Evaluation

The Title I and Title III programs of the Elementary and Secondary Education Act of 1965 provide a comprehensive, timely context for deriving a rationale for educational evaluation. Virtually, every school district in the nation is involved with one or both of these programs. The purposes of these programs respectively are to increase the educational attainment, experiences, and opportunities of disadvantaged children; and to increase the amount and quality of innovation in local education agencies. Both programs are national in scope, design, and broad control. They are coordinated and specifically controlled at the state level and are implemented in local school districts. Together, they provide more than one billion dollars annually to local education agencies.

Figure I contains a conceptualization of the process and decision functions of evaluation as they may exist in federal assistance programs such as the Title I and Title III programs. A set

Figure 1

Feedback Control Loop: Evaluation in Federally Supported Educational Programs



of feedback control loops illustrate the relationships among local, state, and national evaluations of activities of federal assistance programs. In Figure 1, the loop at the right shows local school activities; the center loop, state activities; and the left loop, federal activities. Each loop contains a set of blocks, varied in shape, which represent the major evaluation functions.

Block 1 portrays the local school district's program. This is the local context from which needs for educational change emerge and within which the changes to meet these needs must ultimately occur. It includes the inputs of the system, e.g., the learners, curriculum, staff, organization, policies, finances, physical facilities, and school-community relations, and the outputs of the system, i.e., the cognitive, psychological, physical, and social functioning of its students and alumni.

To the right of Block 1, information collection is depicted by the first segment of curved line. This is a systematic collection at the local level of all information needed for later decisions at local, state, and federal levels.

Block 2 depicts the organization of information. Here, information would be coded according to predetermined categories, processed, e.g., keypunched, filed regularly, and retrieved as needed.

At Block 3, information organized at Block 2 would be analyzed according to decision-making requirements at local, state and national levels and reported to local and state decision-makers.

Block 4 denotes program decisions made at the local level. Local school decision-makers to be served by the evaluation include the Board of Education, the school administration, project supervisors, teachers and principals.

The decisions made at Block 4 would be implemented at Block 5, thus reactivating the cycle with frequent modification of the school program at Block 1. This cycle is continuous.

Returning to Block 3, evaluation reports for the state education department would be prepared annually by all public school districts in the state. At Block 6, the state education department would organize these reports into types of projects and combine information from similar projects. This information would then be analyzed at Block 7 to determine the strengths and weaknesses of the statewide program. The state program officials would use this information to assess the statewide educational needs and problems to make decisions about program emphases and state control at Block 8. Decisions made at Block 8 would be implemented at Block 9, affecting the state program at Block 10, and reactivating the Cycle at Block 1.

At Block 7, annual product evaluation reports from fifty states would be sent to the federal agency. This information would then be organized at Block 11, so that major program thrusts could be examined and analyzed on a nationwide basis at Block 12 and so that reports could be prepared for the Associate Commissioner for

Elementary and Secondary Education, the Commissioner of Education, the Secretary of Health, Education and Welfare, the President, and the Congress. Decisions about program emphases and funding would be made at the federal level at Block 13 and implementation of such decisions at Block 14 would affect the federal program at Block 15, the state program at Block 10, and the local school project at Block 1, thus, reactivating the cycle.

Summarized, Figure 1 demonstrates: (1) information for evaluation at federal, state, and local levels will be collected largely at the local level; (2) this information will form the basis for federal, state, and local decisions which will ultimately affect local operations; and (3) evaluation plans must be developed, communicated, and coordinated at federal, state, and local levels if the information schools provide is to be adequate for assisting in the decision process at each of these levels.

Obviously, to develop an appropriate evaluation system for programs such as Title I and Title III one must first have some knowledge of the decision situations to be served. Optimally, such knowledge of decision situations should answer several questions. First, one should identify the locus of decision-making, in terms of the level(s) at which authority and responsibility for decision-making are vested, e.g., local, state and/or national and within each of these levels. Second, it is desirable to identify the focus of the decisions -- are they related to goals of research, development, training, diffusion, etc.?

Third, one needs knowledge of the substance of the decisions (are they related to mathematics, language arts, etc. and what are the alternatives in each decision situation?). Fourth, one needs to know the function of the decisions--are they for the planning, programing, implementing or recycling of activities? Fifth, one needs knowledge of the objects of the decisions (e.g., persons, places, events, or things?) Sixth, one obviously needs advance knowledge of the timing of decisions. And, finally, one needs knowledge of the relative criticality of decisions.

Considering all of the decision-making variables I have listed above, it is clear that one could identify many, many different kinds of educational decision situation in education. Thus, it would also be possible to identify many different kinds of evaluation. However, it should prove more useful to develop a parsimonious classification system for kinds of educational evaluation which is intermediate between the general conceptual definition of evaluation given above and the many specific applied kinds of evaluation which could be derived from the use of all of the above named variables in a detailed analysis and classification of education decision situations. Then it should be possible to derive useful names for the identified classes of educational evaluation.

To assist in developing a parsimonious classification system for educational decision situations in programs such as Title I and Title III, I have found it useful initially to focus exclusively on

the functions of decisions.[12] I would postulate that functions of decision situations in education may be classified as planning, programing, implementing and recycling. Planning decisions are those which focus needed improvements by specifying the domain, major goals, and specific objectives to be served. Programing decisions specify procedure, personnel, facilities, budget, and time requirements for implementing planned activities. Implementing decisions are those in directing programed activities. And, recycling decisions include terminating, continuing, evolving, or drastically modifying activities.

Four Strategies for Evaluating Educational Programs

Given these four kinds of educational decisions to be served, there are also four kinds of evaluation. These are portrayed in Figure 2 as context, input, process, and product evaluation. Context evaluation would be used when a project is first being planned. Input evaluation would be used immediately after context for specific programing of activities. Process evaluation would be used continuously during the implementation of the project. Product evaluation would most likely be used after a complete cycle of the project. Each of these kinds of evaluation will be considered individually.

Context Evaluation

The major objective of context evaluation is to define the

---

[12]Daniel L. Stufflebeam. "The Use and Abuse of Evaluation in Title III", Theory Into Practice, College of Education, The Ohio State University, Volume VI, Number 3, June 1967.

Figure 2

The CIPP Evaluation Model

A Classification Scheme of Strategies for Evaluating Educational Change

The Strategies

	Context Evaluation	Input Evaluation	Process Evaluation	Product Evaluation
OBJECTIVE	To define the **operation context,** to identify and assess **needs** in the context, and to identify and delineate **problems** underlying the needs.	To identify and assess **system capabilities,** available input **strategies,** and **designs** for implementing the strategies.	To identify or predict, in process, defects in the precedural design or its implementation, and to maintain a record of **procedural events** and **activities.**	To relate **outcome information** to objectives and to context, input, and process information.
METHOD	By describing individually and in relevant perspectives the major subsystems of the context; by comparing actual and intended inputs and outputs of the subsystems; and by analyzing possible causes of discrepancies between actualities and intentions.	By describing and analyzing available human and material resources, solution strategies, and procedural designs for relevance, feasibility and economy in the course of action to be taken.	By monitoring the activity's potential procedural barriers and remaining alert to unanticipated ones.	By defining operationally and measuring criteria associated with the objectives, by comparing these measurements with predetermined standards or comparative bases, and by interpreting the outcome in terms of recorded input and process information.
RELATION TO DECISION-MAKING IN THE CHANGE PROCESS	For deciding upon the **setting** to be served, the **goals** associated with meeting needs and the **objectives** associated with solving problems, i.e., for planning needed changes.	For selecting **sources of support,** solution strategies, and procedural **designs,** i.e., for programming change activities.	For implementing **and refining the program design and procedure,** i.e., for effecting process control.	For deciding to **continue, terminate, modify or refocus** a change activity, and for linking the activity to other major phases of the change process, i.e., for evolving change activities.

environment where change is to occur, the environment's unmet needs, and the problems underlying those needs. For example, the environment may be defined as the inner city elementary schools of a large metropolitan area. Study of such a setting might reveal that the actual reading achievement levels of children in this area are far below what the school system expects for them. This would be the identification of a need, i.e., the context evaluation would have revealed that the children's reading achievement levels need to be raised. As a next step in the context evaluation the school would attempt to identify the reasons for such a need. Are the students receiving adequate instruction? Are the instructional materials appropriate for them? Is there a major language barrier? Is there a high incidence of absenteeism? Is the school's expectation for these students reasonable? Etc. These are what I mean by potential problems. They are potential dilemmas which prevent the achievement of desired goals and thereby result in the existence of needs.

The method of context evaluation begins with a conceptual analysis to identify and define the limits of the domain to be served as well as its major subparts. Next, empirical analyses are performed, using techniques such as sample survey, demography, and standardized testing. The purpose of this part of context evaluation is to identify the discrepancies among intended and actual situations for each of the subparts of the domain of interest and thereby to identify needs. Finally, context evaluation involves both empirical and conceptual analyses, as well as appeal to theory and

authoritative opinion, to aid judgments regarding the basic problems underlying each need.

Decisions served by context evaluation include deciding upon the setting to be served, the goals associated with meeting needs, and the objectives associated with solving problems. Such decisions usually appear in the introductory sections of proposals to funding agencies or in requests for proposals by funding agencies.

Input Evaluation

To determine how to utilize resources to meet program goals and objectives, it is necessary to do an input evaluation. Its objective is to identify and assess relevant capabilities of the proposing agency, strategies which may be appropriate for meeting program goals and designs which may be appropriate for achieving objectives associated with each program goal. The end product of input evaluation is an analysis of alternative procedural designs in terms of potential costs and benefits. Specifically, alternative designs are assessed in terms of their resource, time and budget requirements; their potential procedural barriers; the consequences of not overcoming these barriers; the possibilities and costs of overcoming them; relevance of the designs to program objectives; and overall potential of the design to meet program goals. Essentially, input evaluation provides information for deciding whether outside assistance should be sought for meeting goals and objectives, what strategy should be employed, e.g., the adoption of available solutions or the develop-

ment of new ones, and what design or procedural plan should be employed for implementing the selected strategy.

Methods for input evaluation are lacking in education. The prevalent practices include committee deliberations, appeal to the progessional literature, and the employment of consultants. In a few areas, formal instruments exist to aid decision-makers in making input decisions. In the design of testing programs, one may obtain substantial help by referring to the Buros Mental Measurements Yearbooks.[13] The educational researcher, who wants to select an experimental design, can receive material assistance in identifying and assessing alternative experimental designs by referring to the Campbell-Stanley chapter on experimental design in Gage's Handbook on Research in Teaching.[14] In this chapter, the decision situation posed to the researcher in need of an experimental design is neatly laid out in the form of alternative designs which are relevant to experimental research. Each of these designs is rated regarding its potential to meet criteria of internal and external validity. Further, procedural barriers or sources of invalidity are identified for each of the listed designs.

Decisions based upon input evaluation usually result in the specification of procedures, materials, facilities, schedule, staff

---

[13] Buros, op. cit.

[14] N. L. Gage, Editor, Handbook of Research on Teaching, The American Educational Research Association, Chicago: Rand McNally and Company, 1963.

requirements, and budgets in proposals to funding agencies. From the information provided in the proposals, the funding agencies in turn do an input evaluation to determine whether or not to fund the proposed projects. Funding agencies commonly employ expert consultants to serve as judges in their input evaluations.

Process Evaluation

Once a designed course of action has been approved and implementation of the design has begun, process evaluation is needed to provide periodic feedback to project managers and others responsible for continuous control and refinement of plans and procedures. The objective of process evaluation is to detect or predict, during the implementation stages, defects in the procedural design or its implementation. The overall strategy is to identify and monitor, on a continuous basis, the potential sources of failure in a project. These include interpersonal relationships among staff and students; communication channels; logistics; understandings of and agreement with the intent of the program by persons involved in and affected by it; adequacy of the resources, physical facilities, staff, and time schedule; etc.

As opposed to experimental design evaluation, process evaluation does not require control over assignment of subjects to treatments, nor that the treatments be held constant. Its purpose is to assist project personnel to make their decisions a bit more rational in their continual efforts to improve the quality of the program.

Thus, under process evaluation, the evaluator accepts the program as it is and as it evolves, and monitors the total situation as best he can by focusing the most sensitive and non-intervening data collection devices and techniques that he can obtain on the most crucial aspects of the project. Such evaluation is multivariate, and not all of the important variables can be specified before a project is initiated. The process evaluator focuses his attention on theoretically important variates, but he also remains alert to any unanticipated but significant events. Under process evaluation, information is collected daily, organized systematically, analyzed periodically, e.g., weekly, and reported as often as project personnel require such information, e.g., monthly.

Thus, project decision-makers are not only provided with information needed for anticipating and overcoming procedural difficulties, but also with a record of process information to be used later for interpreting project outcomes.

Product Evaluation

Product evaluation is used to determine the effectiveness of the project after it has run full cycle. Its objective is to relate outcomes to objectives and to context, input, and process, i.e., to measure and interpret outcomes.

The method is to operationally define and measure criteria associated with the objectives of the activity, to compare these measurements with predetermined absolute or relative standards, and

to make rational interpretations of the outcomes using the recorded context, input, and process information. Criteria for product evaluation may be either instrumental or consequential, a distinction pointed out earlier by Scriven.[15] Instrumental criteria are related to program outcomes which contribute to the achievement of behavioral objectives. Clark and Guba have developed a taxonomy of instrumental objectives and associated criteria which are related to educational change.[16] An adaptation of their scheme is presented as Figure 3. Consequential criteria are primarily those pertaining to behavioral objectives. Bloom's Taxonomy of Educational Objectives[17] is useful in the identification of consequential objectives.

In the change process, product evaluation provides information for deciding to continue, terminate, modify or refocus a change activity, and for linking the activity to other phases of the change process. For example, a product evaluation of a program to develop after school study for students from disadvantaged homes might show that the development objectives have been satisfactorily achieved

---

[15]Michael Scriven. The Methodology of Evaluation, Bloomington, Indiana: Indiana University, Social Science Education Consortium, Publication #110, 1965.

[16]David L. Clark and Egon G. Guba. "An Examination of Potential Change Roles in Education," Paper read at the Symposium on Innovation in Planning School Curricula, Airlie House, Virginia, October, 1965.

[17]Benjamin S. Bloom. Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain, New York: Longmans, Green and Company, Inc., 1956.

A Process Chart Depicting the Role of Evaluation in the Change Process*

	AGENCY	OBJECTIVE	PROCESS	CRITERIA	RELATION TO CHANGE
RESEARCH	Universities, Research and Development Institutions, and Regional Laboratories.	To advance knowledge, i.e., to depict, correlate, conceptualize, and test.	RESEARCH → EVALUATION	Validity (internal and external).	Provides basis for invention.
DEVELOPMENT	Universities, Research and Development Institutions, Regional Laboratories, and Industries.	To formulate a new solution to an operating problem or to a class of operating problems i.e., to innovate.	INVENTION → EVALUATION	Face validity (appropriateness); estimated viability; impact (relative contribution).	Produces the invention.
		To draft a plan for constructing the innovation, i.e., to construct the blue-print.	DESIGN → EVALUATION	Feasibility (production and utilization); tractability (ease of managing, controlling, and instructing in the use of).	Engineers the invention to fit the characteristics of the target situation.
		To build the components, i.e., to construct.	CONSTRUCTION → EVALUATION	Design specifications; individual performance.	Produces the components necessary for implementing the design.
		To integrate the components into an operating system, i.e., to finalize for marketing.	ASSEMBLY → EVALUATION	Design specifications; total performance, viability, efficiency.	Produces the operationalized package for use.
DIFFUSION	Government, Universities, and Regional Laboratories.	To create widespread awareness of the invention among practitioners, i.e., to inform.	DISSEMINATION → EVALUATION	Intelligibility; fidelity; pervasiveness; impact (extent to which it affects key targets).	Informs about the invention.
		To afford an opportunity to examine and assess operating qualities of the invention, i.e., to build conviction.	DEMONSTRATION → EVALUATION	Credibility; convenience; evidential assessment.	Builds confidence about the invention.

		To integrate the components into an operating system, i.e., to finalize for marketing.	ASSEMBLY → EVALUATION	Design specifications: total performance, viability; efficiency.	Produces the coordinated operating system.
DIFFUSION	Government Universities, and Regional Laboratories.	To create widespread awareness of the invention among practitioners, i.e., to inform.	DISSEMINATION → EVALUATION	Intelligibility; fidelity; pervasiveness; impact (extent to which it affects key targets).	Informs about the invention.
		To afford an opportunity to examine and assess operating qualities of the invention, i.e., to build conviction.	DEMONSTRATION → EVALUATION	Credibility; convenience; evidential assessment.	Builds knowledge about the invention.
ADOPTION	Universities, Regional Laboratories, and Schools.	To train local personnel to manage, operate, service, and and utilize the innovation, i.e., to staff.	TRAINING → EVALUATION	Quantity, continuity, aptitudes, motivation and proficiency of trained personnel.	Establishes and maintains viability for operating the innovation.
		To build familiarity with the invention and provide a basis for assessing the quality, value, fit, and utility of the invention in a particular institution, i.e., to test.	TRIAL → EVALUATION	Adaptability; feasibility; action.	Tries out the invention in the context of a particular situation.
		To fit the characteristics of the invention to the characteristics of the adopting institution, i.e., to operationalize.	INSTALLATION → EVALUATION	Effectiveness, efficiency.	Operationalizes the invention for use in a specific institution.
		To assimilate the invention as an integral and accepted component of the system, i.e., to establish.	INSTITUTIONALIZATION → EVALUATION	Continuity; valuation; support.	Establishes the invention as a part of an ongoing program; converts it to an innovation.

* Based Upon "A Classification Scheme of Processes Related to and Necessary for Change in Education" by David L. Clark and Egon G. Guba

and that the developed innovation is ready to be diffused to other schools which need such an innovation.

Given these four kinds of evaluation it is next necessary to consider methodology for implementing them. This problem is considered in the next section of this paper.

The Structure of Evaluation Design

Once an evaluator has selected an evaluation strategy, e.g., context, input, process, or product, he must next select or develop a design to implement his evaluation. This is a difficult task since few generalized evaluation designs exist which are adequate to meet emergent needs for evaluation. Thus, educators must typically develop evaluation designs de novo. The remainder of this paper is an attempt to provide a general guide for developing evaluation designs. Specifically, I will attempt to define design in general terms and to explicate the general structure of designs for educational evaluation. Hopefully, this general treatment of evaluation design will be of some help to educators in ordering their minds as they approach problems of designing evaluations. Also, I am hopeful that the following material might stimulate methodologists who are more capable than I to develop generalized designs for context, input, process, and product evaluation.

Design Defined

In general, design is the preparation of a set of decision situations for implementation toward the achievement of specified

objectives. This definition says three things. First, one must identify the objectives to be achieved through implementation of the design. In a product evaluation, for example, such an objective might be to make a determination of whether all students in a remedial reading program attained specified levels of specific reading skills. Second, this definition says that one should identify and define the decision situations in the procedure for achieving the evaluation objective. For example, in the remedial reading case cited above one would want to identify the available measuring devices which might be appropriate for assessing the specified reading skills. Third, for each identified decision situation the evaluator needs to make a choice among the available alternatives. Thus, the completed evaluation design would contain a set of decisions as to how the evaluation is to be conducted and what instruments will be used.

It should be useful to evaluators to have available a list of the decision situations which are common to many evaluation designs. This would enable them to approach problems of evaluation design in a systematic manner. Further, such a list could serve as an outline for the content of evaluation sections in research and development proposals. Funding agencies should also find such a list useful in structuring their general guidelines for evaluations which they provide to potential proposal writers. Also, such a list should be useful to training agencies for defining the role of the evaluation specialist.

Figure 4 is an attempt to provide such a general list of decision situations for evaluation designs. By presenting this general list I am asserting that the structure of evaluation design is the same for context, input, process, or product evaluation. This structure includes six major parts. These are 1) focusing the evaluation, 2) information collection, 3) information organization, 4) information analysis, 5) information reporting, and 6) the administration of evaluation. Each of these parts will be considered separately.

Focusing the Evaluation

The first part of the structure of evaluation design is that of focusing the evaluation. The purpose of this part is to spell out the ends for the evaluation and to define policies within which the evaluation must be conducted. Specifically, this part of evaluation design includes four steps.

The first step is to identify the major levels of decision-making for which evaluation information must be provided. For example, in the Title III program of the Elementary and Secondary Education Act evaluative information from local schools is needed at local, state and national levels. It is important to take all relevant levels into account in the design of evaluations since different levels may have different information requirements and since the different agencies may need information at different times.

Having identified the major levels of decision-making to be served by evaluation, the next step is to identify and define the

Figure 4

Daniel L. Stufflebeam
January 1968

DEVELOPING EVALUATION DESIGNS

The logical structure of evaluation design is the same for all types of evaluation, whether context, input, process or product evaluation. The parts, briefly, are as follows:

A. Focusing the Evaluation
   1. Identify the major level(s) of decision-making to be served, e.g., local, state, or national.
   2. For each level of decision-making, project the decision situations to be served and describe each one in terms of its locus, focus, criticality, timing, and composition of alternatives.
   3. Define criteria for each decision situation by specifying variables for measurement and standards for use in the judgment of alternatives.
   4. Define policies within which the evaluation must operate.

B. Collection of Information
   1. Specify the source of the information to be collected.
   2. Specify the instruments and methods for collecting the needed information.
   3. Specify the sampling procedure to be employed.
   4. Specify the conditions and schedule for information collection.

C. Organization of Information
   1. Provide a format for the information which is to be collected.
   2. Designate a means for coding, organizing, storing, and retrieving information.

D. Analysis of Information
   1. Select the analytical procedures to be employed.
   2. Designate a means for performing the analysis.

E. Reporting of Information
   1. Define the audiences for the evaluation reports.
   2. Specify means for providing information to the audiences.
   3. Specify the format for evaluation reports and/or reporting sessions.
   4. Schedule the reporting of information.

F. Administration of the Evaluation
   1. Summarize the evaluation schedule.
   2. Define staff and resource requirements and plans for meeting these requirements.
   3. Specify means for meeting policy requirements for conduct of the evaluation.
   4. Evaluate the potential of the evaluation design for providing information which is valid, reliable, credible, timely, and pervasive.
   5. Specify and schedule means for periodic updating of the evaluation design.
   6. Provide a budget for the total evaluation program.

decision situations to be served at each level. Given our present low state of knowledge about decision-making in education, this is a very difficult task. However, it is also a very important one and should be done as well as is practicable. First, decision situations should be identified in terms of those responsible for making the decisions, e.g., teacher, principals, the board of education members, state legislators, etc. Next, major types of decision situations should be identified, e.g., appropriational, allocational, approval, or continuation. Then these types of decision situations should be classified by focus, e.g., research, development, diffusion or adoption in the case of instrumental outcomes, or knowledge or understanding in the case of consequential outcomes. (This step is especially helpful toward identifying relevant evaluative criteria.) These identified decision situations should then be analyzed in terms of their relative criticality. In this way relatively less important decisions which would expend evaluation resources needlessly can be eliminated from further consideration. Next, the timing of the decision situation to be served should be estimated so that the evaluation can be geared to provide relevant data prior to the time when decisions must be made. And, finally, an attempt should be made to explicate each important decision situation in terms of the alternatives which may reasonably be considered in reaching the decision.

Once the decision situations to be served have been explicated, the next step is to define relevant information requirements. Specifically, one should define criteria for each decision situation by specifying variables for measurement and standards for use in the

judgment of alternatives.

The final step in focusing the evaluation is to define policies within which the evaluation must operate. For example, one should determine whether a "self evaluation" or "outside evaluation" is needed. Also, it is necessary to determine who will receive evaluation reports and who will have access to them. Finally, it is necessary to define the limits of access to data for the evaluation team.

Collection of Information

The second major part of the structure of evaluation design is that of planning the collection of information. This section must obviously be keyed very closely to the criteria which were identified in the Evaluation Focus part of the design.

Using those criteria one should first identify the sources of the information to be collected. These information sources should be defined in two respects; first, the origins for the information, e.g., students, teachers, principals or parents, and second, the present state of the information, i.e., in recorded or non-recorded form.

Next, one should specify instruments and methods for collecting the needed information. Examples include achievement tests, interview schedules and searches through the professional literature.

Michael and Metfessel[18] have recently provided a comprehensive list of instruments with potential relevance for data collection in evaluations.

For each instrument that is to be administered, one should next specify the sampling procedure to be employed. Where possible, one should avoid administering too many instruments to the same person. Thus, sampling without replacement across instruments can be a useful technique. Also, where total test scores are not needed for each student, one might profitably use multiple matrix sampling where no student attempts more than a sample of the items in a test.

Finally, one should develop a master schedule for the collection of information. This schedule should detail the interrelations between samples, instruments, and dates for the collection of information.

Organization of Information

A frequent disclaimer in evaluation reports is that resources were inadequate to allow for processing all of the pertinent data. If this problem is not to arise, one should make definite plans regarding the third part of evaluation design: Organization of

---

[18]Newton S. Metfessel and William B. Michael. "A Paradigm Involving Multiple Criterion Measures for the Evaluation of the Effectiveness of School Programs," Educational and Psychological Measurement, 1967, 27, 931-936.

Information. Organizing the information that is to be collected includes providing a format for classifying information and designating means for coding, organizing, storing, and retrieving the information.

Analysis of Information

The fourth major part of evaluation design is analysis of information. The purpose of this part is to provide for the descriptive or statistical analyses of the information which is to be reported to decision-makers. This part also includes interpretations and recommendations. As with the organization of information it is important that the evaluation design specify means for performing the analyses. The role should be assigned specifically to a qualified member of the evaluation team or to an agency which specializes in doing data analyses. Also, it is important that those who will be responsible for the analysis of information participate in designing the analysis procedures.

Reporting of Information

The fifth part of evaluation design is the reporting of information. The purpose of this part of a design is to insure that decision-makers will have timely access to the information they need and that they will receive it in a manner and form which facilitates their use of the information. In accordance with the policy for the evaluation, audiences for evaluation reports should be identified and defined. Then means should be defined for providing information to each audience. Subsequently, the format for evaluation

reports and reporting sessions should be specified. And, finally, a master schedule of evaluation reporting should be provided. This schedule should define the interrelations between audiences, reports, and dates for reporting information.

Administration of Evaluation

The last part of evaluation design is that of administration of the evaluation. The purpose of this part is to provide an overall plan for executing the evaluation design. The first step is to define the overall evaluation schedule. For this purpose it often would be useful to employ a scheduling technique such as Program Evaluation and Review Technique. The second step is to define staff requirements and plans for meeting these requirements. The third step is to specify means for meeting policy requirements for conduct of the evaluation. The fourth step is to evaluate the potential of the evaluation design for providing information which is valid, reliable, credible, timely, and pervasive. The fifth step is to specify and schedule means for periodic updating of the evaluation design. And, the sixth and final step is to provide a budget for the evaluation.

Finally, I have reached the end of my paper. While I have only scratched the surface regarding educational evaluations, it is clear to me that the design and analysis of educational evaluation is a most complex and difficult undertaking. Surely, all of us who are committed to reshaping the world of educational evaluation must work very, very hard if we are to make any progress. If progress is not

made in this area, I am convinced that education will be a casualty for want of adequate information to support vital decisions in and about education.